

Live URLs: Breathing life into URLs

Natarajan Kannan
Infosys Technologies Limited.
Bangalore
India

natarajan_kannan@infosys.com

Toufeeq Hussain
Infosys Technologies Limited.
Bangalore
India

toufeeq_hussain@infosys.com

ABSTRACT

This paper provides a novel approach to use URI fragment identifiers to enable HTTP clients to address and process content, independent of its original representation.

Categories and Subject Descriptors

H.4.3 [Internet]; I.7.1 [WEB]; I.7.2 [HTML, XML, XHTML]

General Terms

Standardization, Experimentation

Keywords

ACM proceedings, HTTP, HTML, URL, browsers, fragment identifier, web content, web addressing

1. INTRODUCTION

Since the inception of the World Wide Web, URI/URLs [1] have been the most common way of locating HTTP resources. Today, people exchange information on the location of web based resources using URLs. These URLs could include a fragment identifier [2] (if any) so that the desired information can be accessed quickly. Till today, the fragment identifier's interpretation has been to denote pre-defined sections of a web page by its author. We propose to use the information in the fragment identifier for arbitrary content location, presentation and processing through enhanced HTTP clients. We refer to URLs, which use the fragment identifier for these content-centric purposes, as 'Live URLs'.

2. BENEFITS

A 'Live URL' enabled client has the capability to address content on any HTML page arbitrarily. The creation and usage of 'Live URLs' is independent of the page creator and the underlying HTML source. 'Live URLs' can be used in the same manner as existing URLs. Addressing and client side processing of web content is not restricted by a central server or entity. Web annotation projects like W3C's Annotea [3] can take advantage of addressing and processing capabilities of 'Live URLs'.

3. LIVE URLs

The purpose of the fragment identifier as per the W3C URL addressing standards is to represent a fragment or a sub function within an object [2] [1]. The interpretation of the fragment identifier is dependent on the media type of the retrieval result. The current interpretation of the fragment identifier for HTML pages is to locate pre-defined anchor references in the HTML source. Though the URL addressing semantics mention that specific syntaxes like line and character range, or coordinates can be used as the fragment identifier, we feel it has been grossly under-utilized. The following example illustrates how the fragment identifier can be effectively used.

3.1 Example

A typical URL with a fragment identifier would be: `http://www.example.com/url.html#urgent`. In this case the browser focuses on the section which is anchored by name `urgent`.

A 'Live URL' would look like: `http://www.example.com/url.html#top=13,left=46,right=89,bottom=02,action=highlight` where the co-ordinates specify which section is to be prominently displayed. This scheme also gives the flexibility of how we want to process and display the specified content.

We refer to the new URLs as 'Live URLs' as they bring dynamism to otherwise static content.

3.2 Applications

Some of the applications of 'Live URLs' are discussed below.

3.2.1 Locating content

'Live URLs' can be used to specify information about the location of content within a web page. The location identifier could be one of the following:

DOM elements and offsets within them Using unique DOM (Document Object Model) elements and offsets to identify the starting and ending points of selected content, information about the selected content could be specified in the 'Live URL'. The Ahoy project [4] has an implementation which generates unique DOM identifiers for arbitrary points of a HTML document.

Search strings The 'Live URL' could also contain "search strings" which could be used to locate sections of a web page.

XPointer 'Live URLs' could also contain XPointers (XML Pointer Language) [5], which are used to locate content in XML documents.

Byte offsets The HTTP protocol [6] specifies that a byte range can be used to denote a subset of a larger HTTP entity. Byte range specifications in HTTP refer to a sequence of bytes in the entity body. The byte range which is required for display of content can be passed as part of the 'Live URL'.

Using web-page co-ordinates This method is based on the fact that the whole web page is mapped onto a co-ordinate based grid. The starting and ending co-ordinates of a selected portion of a web page are specified in the 'Live URL' as name-value pairs.

3.2.2 Dynamic content processing

In addition to locating content, 'Live URLs' can also specify a host of content specific actions like highlighting, searching, zooming, removal etc.

3.2.3 Client side scripting

'Live URLs' can also be used as a generic mechanism of providing input parameters to client-side scripts. Client side scripting technologies like AHAH [7] can benefit from 'Live URLs' as the remote HTML/XHTML fragments, can be addressed using 'Live URLs'.

3.3 HTTP Clients

The capability to handle 'Live URLs' will be built into HTTP clients (primarily web browsers). As per current implementations, clients retrieve the document specified by the given URL and render it. If there is a fragment identifier specified in the URL, the focus is shifted to that section. In this case, the client is enhanced to interpret the fragment information that is part of the 'Live URL' and suitably render the referenced content. Also, existing clients will not reject 'Live URLs' totally, but show the entire web page without any modifications. This stems from the fact that, non-existent fragment identifiers are discarded by clients.

We are currently working on a reference implementation as an extension [8] to the Mozilla Firefox web browser [9].

4. USE CASES

Possible use cases of the proposed 'Live URL' concept are discussed below.

4.1 Client use case

A user while reading a web page, comes across a section of the page which he wants to share. He selects the section and obtains a 'Live URL' for the selected portion from a browser menu option. He then passes on this 'Live URL' to a receiver. The receiver opens the 'Live URL' using his 'Live URL'-aware browser. The 'Live URL'-aware browser retrieves the document and uses the information in the 'Live URL' to display the content that was selected by the sender in a prominent manner.

4.2 Server/Web application use case

A typical example of a server/web application use case would be a search engine which returns 'Live URLs' as part of search results. The user can open the 'Live URLs' to view the relevant content directly.

Another use case would be a situation, where the server would want to offload some of the processing to the client. For example, when the server wants to highlight certain content on a web page, instead of generating the highlighted content on the server, it could redirect the client to a 'Live URL', which has necessary inputs for the client to perform the highlighting.

5. FUTURE WORK

Future work on 'Live URLs' could include development of client side scripting frameworks which could take advantage of parameters passed as part of 'Live URLs'. It could also include enhancing HTTP servers to support generation of 'Live URLs' based on client's capabilities.

6. ACKNOWLEDGEMENTS

The authors thank Rajesh Balakrishnan, Naveen Krishnan Unni, Raghuv eer, Mahendra, Sanjay Nayak and Basavesh of Infosys Technologies Ltd. for their encouragement, support and ideas.

7. CONCLUSION

In this paper, we have proposed a novel approach of using URI fragment identifiers to enable HTTP clients to address and process content, independent of its original representation. We foresee a variety of applications of the concept proposed, ranging from simple content sharing to accurate and focussed search results. The proposed concept can be extended used to selectively access audio/video content. We believe fragment identifiers can be used to make HTTP clients smarter and 'Live URLs' are a step in that direction.

8. REFERENCES

- [1] T Berners-Lee, R Fielding, and L Masinter. Uniform resource identifiers (uri): Generic syntax. <http://www.ietf.org/rfc/rfc2396.txt>.
- [2] W3C. Fragment-id of an uri. http://www.w3.org/Addressing/URL/4_2_Fragments.html.
- [3] W3C. Annotea project. <http://www.w3.org/2001/Annotea>.
- [4] Brian Donovan. Everything can be a link with mozilla's window.getSelection(). <http://dev.lophty.com/ahoy/index.htm>.
- [5] W3C. Xpointer framework. <http://www.w3.org/TR/xptr-framework/>.
- [6] R Fielding, J Gettys, J Mogul, H Frystyk, L Masinter, P Leach, and T Berners-Lee. Hypertext transfer protocol – http/1.1. <http://www.ietf.org/rfc/rfc2616.txt>.
- [7] Asynchronous html and http. <http://microformats.org/wiki/rest/ahah>.
- [8] Natarajan Kannan and Toufeeq Hussain. Liveurls. <http://liveurls.mozdev.org>.
- [9] Firefox. <http://www.mozilla.com/firefox/>.