

A Simple Reference Management System for HTML

L. K. Manickum

lelanthran@gmail.com

April 16, 2013

Abstract

Correct management of references is an important part of academic writing. There exists tools to help manage the list of references cited that work with current word-processing and other academic authoring software such as L^AT_EX. In this paper I present a modest and simple proposal together with a prototype outlining the creation of a similar tool for HTML.

Keywords: *HTML, references database, citation management, bibliography management, online authoring*

1 Introduction

In this article I use the word “*citation*” and “*reference*”. I use “*citation*” to refer to the common shorthand reference within an article text[6] that denotes an entry in the references list, or bibliography. An example would be as follows (citation is emphasised in bolded font):

The quick brown dog juggled**[1]** over the lazy fox**[2]**.

The “[1]” and “[2]” in the above passage is what I mean when I use the word “*citation*” in this article. I use the word “*reference*” to refer to an entry in the bibliography at the end of the article. For example, each of the following is a reference:

[1] **Playing with Pins**, Pennywise Clown, Feb 13, 1982, Laughter press.

[2] **Saving the Henhouse**, Old McDonald, Dec 2, 1999, Cornfield Press.

Each number in the bibliography corresponds to one or more numbered citations within the article text. Each numbered (or otherwise identified) citation will correspond to one, and only one, reference in the bibliography.

Academic papers are rarely, if ever, a standalone piece of work. Authors build upon previously published ideas and concepts to generate new ideas and concepts. Thus a critical part of any serious academic paper is the list of references that the paper cites in support of the argument or thesis. Popular word-processing software and other authoring software provides tools to enable the proper management of the list of references and to ensure, amongst other things, that reference styles are

adhered to and that citations are correct, and that no superfluous publications are listed in the bibliography.

Many of the tools that exist to facilitate the creation and maintenance of a *References Database* work together with the authoring tool to ensure that the article correctly and accurately cites all references, and that all references are presented in a bibliography at the end of the article. None of the 28 popular reference management tools[10] integrate with HTML in the seamless manner that I propose below. While it is true that many authoring tools can export to HTML, this is useless when editing a blog post, or a Wiki. HTML is simply not considered an important enough medium for academic publications that it deserves its own reference management tools.

Due to the growing importance of HTML[7] as a content delivery mechanism, and in some cases as the main content delivery mechanism[11][3] even in areas that are non-ICT related[1], many authors are turning to HTML as a medium for delivering their ideas to the world[2][5], or even just for private consumption within an organisation[4]. A relatively new phenomenon is the weblog, or “blog” as it is more commonly known. These blogs are written entirely in HTML. However, other than a `<cite>` tag[9] for phrase elements and a `cite` attribute for the `<blockquote>` tag[8], there is absolutely no support for reference or citation management in the HTML specification. The result are blog articles with incorrect citation, mentioned (but missing) references or worse, no references at all. The quality of the blog writing or the feasibility of the authors idea is difficult to gauge when the material has no supporting references.

It is hard to fault the authors themselves when it is known that adding citations to HTML is cumbersome, tedious and error-prone. Manually adding in the required HTML code tags to provide citations and a bibliography also results in inconsistent referencing and citation styles, sometimes *even within a single article!* To alleviate the situation I looked into the feasibility of providing a References Database management tool that will work seamlessly within the HTML specification and provide a layer of ease, accuracy and consistency for both the author and the reader.

2 Manually Adding Citations and a Reference List in HTML

Hypertext was always seen as possessing the potential to be the ultimate authoring tool, in particular providing facilities for citation and bibliographic reference management. In printed publications, the reader of an article, upon seeing a citation in the material, will immediately refer to the bibliography at the end of the article to learn the Title, Author, Publisher, Date, etc of the citation. Once learning these salient facts about the cited work, the reader will then go into a library and physically find the cited work.

In contrast, hypertext (in theory) allows the reader of an article, upon coming across a citation in the middle of the article text, to simply follow that citation to see the item in the bibliography at the end of the article. Upon seeing the bibliography entry for that citation containing the Title, Author, Date, etc, the reader can then simply follow *that* text to the full article. In practice this turned out not to be the case. The most popular hypertext system, HTML over HTTP across the World-Wide-Web, doesn't natively allow such a thing.

Adding a citation plus the correct corresponding entry in the bibliography when writing in HTML is a long process with much simple calculation and checking needing to be done by the author. For example, to create the following cited material (and corresponding references in the bibliography), the author needs to perform the following steps:

1. Add in the HTML code to display the citation, such as:

```
<p>The quick brown dog juggled<a href="#entry_1">[1]</a> over  
the lazy fox<a href="#entry_2">[2]</a>.
```

</p>

2. Edit the bibliography at the end of the article to ensure that the references exist in the correct order and with the correct names so that clicking the link for citation “[1]” will take the reader to the correct entry in the bibliography. This is done as follows:

```
<a name="entry_1">
  [1] Playing with Pins,
      Pennywise Clowen, Feb 13, 1982,
      Laughter press.</a>

<a name="entry_2">
  [2] Saving the Henhouse
      Old McDonald, Dec 2, 1999
      Cornfield press.</a>
```

This results in the correct citation and accurate reference.

Should the author then add in another citation to the article, for example, in order to provide a citation for “quick” in the already written sentence, then the author would have to do the following:

1. Change the numbered citations in the article text to the following:

```
<p>The quick<a href="#entry_1">[1]</a> brown dog
  juggled<a href="#entry_2">[2]</a>
  over the lazy fox<a href="#entry_3">[3]</a>.
</p>
```

2. Change the bibliography to reflect the new numbering above:

```
<a name="entry_1">
  [1] Definitive Dictionary of Easy Words,
      M. Python, Apr 1, 1979,
      Flying Circus press.</a>

<a name="entry_2">
  [1] Playing with Pins,
      Pennywise Clowen, Feb 13, 1982,
      Laughter press.</a>

<a name="entry_3">
  [1] Saving the Henhouse
      Old McDonald, Dec 2, 1999
      Cornfield press.</a>
```

As can be seen, this is a time-consuming process that is particularly well-suited to generating non-obvious but ultimately disastrous errors. For example, the author may forget to put a reference (that a citation uses) into the bibliography, or the citation number for “*quick*” may change due to a new citation added prior to it, and the author forgets to change *all* the other citation numbers in the article, leading to two different references having the same citation number. In another case, the citation numbers may all be amended correctly in the article, but the author may have mixed up the order of references in the bibliography, leading to incorrect citations. In fact, there are so many ways that this could go wrong via simple errors that it is actually easier to specify when it will be correct. It will only be correct in the event that the author studiously re-reads the entire article in a thorough fashion *every single time a citation is added!*

It is no small wonder, then, that authors using the medium of HTML over HTTP on the WWW prefer to simply forgo the entire process of citing their arguments and listing their references. It is quite possible that it would make the citation of any argument a long and thoroughly boring exercise, said exercise having little chance of success anyway.

3 Automating the Procedure of HTML Citation and Reference Management

Using the built-in scripting capability of HTML in the form of ECMAScript¹, more commonly known as *Javascript*, one can easily automate the tasks above. The steps that can be fully scripted to work within the web browser are:

1. Add references to an array that will function as the *References Database*.
2. Go through the article and note, for each citation, which number it is. For example, the first citation we get should be [1], the second citation should be [2], etc. If we come across citation [1] again, we should not change its number, for then the order will be lost.
3. At the point of each citation, we print out the citation number and make that printed text a hypertext link to the corresponding entry in the bibliography.
4. At the end of the article, print out a list of references in the order they were cited and with the correct numbering. We also ensure that each reference in the bibliography has the correct name so that clicking upon a citation in the article will result in the browser scrolling to the correct entry in the bibliography.
5. For each entry in the bibliography that is printed out, if the source is available via a url, then make that bibliographic entry a hypertext link to the source material.

I built a prototype to do all of the above, and the code snippets below show how it is used in normal usage. The example above is used to show the increase in brevity and readability.

1. A function exists called `addRef()` that will create an object with the given field/value pairs as a single reference and add that object to an array of all references. The function is designed so that the caller (author) can simply call the function with all the fields that are necessary, for example the Title, Author, Date, etc, and the function will store a single object with all the fields given. The fields can be specified in any order, or even simply omitted completely. The example usage is:

```
addRef ({name   : " juggler_pins",
        author  : "Pennywise Clown",
        day     : "13",
        month   : "Feb",
        year    : "1984",
        volume  : "12",
        journal : "Quantum psychotics",
        title   : "Playing with Pins",
        publisher : "Laughter press"
        });

addRef ({name : "fox_behaviour",
        title  : "Saving the Henhouse",
```

¹ECMA-262

```

author : "Old McDonald",
year : "1999",
month : "Dec",
issue : "17",
publisher: "Cornfield Press"
});

```

As can be seen, each entry has to have a unique mandatory name assigned by the author of the article. This is in line with the way many of the popular reference management software currently works and users are both aware of this need and are used to this requirement. Each entry does not need all fields filled in, and the fields can be specified in any order. The design is such that the author can simply keep all references in a single file and then pull in the references in the HTML article simply by doing: `<script src="references.js"></script>`

2. A non-standard tag called `cit` is recognised; the element data contained by this tag is used as the reference argument specifying which of the references is being cited. The argument is the unique name given in `addRef()` above. More than one argument can be specified within the `cit` tag. This tag is simply replaced by the reference number in the form of `[?]` where the ‘?’ is a number denoting the citations entry in the reference list. It then links this text to the actual entry in the bibliography so that clicking on the citation will take the reader to the correct entry in the bibliography.²

The example usage is:

```

<p>The quick brown dog
    juggled<cit>juggler_pins</cit>
    over the lazy fox<cit>fox_behaviour</cit>
</p>
...
<h3>Bibliography</h3>
<bibliography> </bibliography>

```

3. A function called `jsNote()` exists, and this function, called without any arguments at all, simply prints out the references made by all the citations in the article in the correct order that they were made, and serves as a placeholder for the citations to link to. As an example, at the end of the article the author would do:

```

<script> jsNote () </script>

```

The author need not worry about forgetting to add a reference that has a citation number, the javascript functions will ensure that the numbers all match up correctly. The author also need not care that the references in the bibliography are presented in the correct order, the software ensures that it displays in the correct order. In fact, all the author has to do is simply use the two javascript functions and all the work is taken care of by the javascript. In the event that a new citation is added before any of the existing citations the software will calculate the correct numbering of the remaining citations as well as ensure that the corresponding references in the bibliography are positioned and numbered correctly.

The javascript code can be retrieved from <http://www.lelanthran.com/apps/jsnote.js> and an example references database can be retrieved from

²More than one bibliography can be inserted; however this is rarely productive. While it usually makes sense to place the bibliography at the bottom of a document, the “bibliography” tag is supported anywhere, not just at the end of the document.

<http://www.lelanthran.com/apps/references.js>.

The code is fully documented for the user and copious examples are provided.

4 Conclusion and Further Avenues of Development and Research

The user of this functionality has to be fully aware that they are using a non-standard tag (“cit”). A further development not fully discussed in this article is the `setCiteTag()` and `setBibTag` functions, which change the default tags used to identify citations and bibliographies from the default values of “cit” and “bibliography”, respectively.

There are of course limitations in the current prototype incarnation of this software. For example, the actual styles (Harvard, MLA, etc) should be controllable by the author, such as having a single function called `setStyle()` which would set the style to whatever is specified and then use that style throughout the article. A further enhancement could perhaps allow for something similar to footnotes in printed material. By simply having a citation text of ‘*’ that, on a click, takes one to the relevant footnote, the author can add extra non-reference information to the article in the form of footnotes.

Further revisions of this software would encompass the footnote functionality. I look forward to user-provided ideas.

5 Downloads

Javascript code to include: <http://www.lelanthran.com/apps/jsnote.js>

Example references file: <http://www.lelanthran.com/apps/references.js>

6 References

References

- [1] Maged Boulos, Inocencio Maramba, and Steve Wheeler. Wikis, blogs and podcasts: a new generation of web-based tools for virtual collaborative clinical practice and education. *BMC Medical Education*, 6(1):41, 2006.
- [2] Matthew D. and Barton. The future of rational-critical debate in online public spheres. *Computers and Composition*, 22(2):177 – 190, 2005.
- [3] Daneial W Drezner and Henry Farrel. The power and politics of blogs. Springer Science+Business Media, September 2007.
- [4] Darlene Fichter. Intranets, wikis, blikis and collaborative working. *Online*, 29(5):47–50, September/October 2005.
- [5] Thomas P. Mackey. The social informatics of blog and wiki communities: Authoring communities. Proceedings of Canadian Association for Information Science, 2007.
- [6] College of Arts and Sciences. Recognizing and avoiding plagiarism. Technical report, Cornell University, 2005.
- [7] L.G. Roberts. Beyond moores law: Internet growth trends. *Computer*, 33(1):117,118,119, January 2000.

- [8] W3C. Html blockquote tag. Technical report, W3C, 2012. http://www.w3schools.com/tags/tag_blockquote.asp.
- [9] W3C. Html em strong dfn code samp kbd var cite tags. Technical report, W3C, 2012. http://www.w3schools.com/tags/tag_phrase_elements.asp.
- [10] Wikipedia. Comparison of reference management software. Technical report, Wikipedia, 2012. http://en.wikipedia.org/wiki/Comparison_of_reference_management_software.
- [11] Jeremy B. Williams and Joanne S. Jacobs. Exploring the use of blogs as learning spaces in the higher education sector. *Australasian Journal of Educational Technology*, 20(2):232–247, 2004. The contents of this journal can be freely accessed online via the journal’s web page (see hypertext link).